# Final Report

## Summary of Key Findings

### Predictive model performance

- Accuracy:
  - The ANN model achieved an overall accuracy of 79.18%, meaning that approximately 79% of the predictions (churners and non-churners) were correct.
- Precision (Churn Class):
  - Precision for the churn class was 64.71%, indicating that when the model predicted churn, it was correct about 65% of the time.
- Recall (Churn Class):
  - Recall for the churn class was 47.70%, meaning the model identified only approximately 48% of actual churners.
- F1-Score:
  - The F1-score for churn was 54.91%, reflecting the balance between precision and recall.
- Limitations:
  - The model struggled with false negatives (missed churners), largely due to the class imbalance in the dataset.

### Key factors contributing to churn

**Customer tenure:**

- Observation:
  - Customers with shorter tenure (less than one year) exhibited higher churn rates.
- Insight:
  - Early-stage customers are likely to churn due to dissatisfaction, lack of engagement, or inadequate onboarding.

**Monthly charges:**

- Observation:
  - High monthly charges were strongly correlated with higher churn rates.
- Insight:
  - Customers on premium plans are more likely to leave if they perceive a lack of value or face financial strain.

**Contract type:**

- Observation:
  - Customers on Month-to-Month contracts had the highest churn rates, whereas those on One-Year or Two-Year contracts churned significantly less.
- Insight:

  - Month-to-Month customers have more flexibility to leave, making them more likely to churn.

**Internet service type:**

- Observation:
  - Customers using Fiber Optic internet had higher churn rates compared to those on DSL or other services.
- Insight:
  - High costs or unmet performance expectations may drive Fiber Optic customers to churn.

**Senior citizens:**

- Observation:
  - Senior citizens churned at higher rates than non-senior customers.
- Insight:
  - Cost sensitivity and potential challenges in service adoption may contribute to this trend.

## Identification of Factors Contributing to Churn and Retention

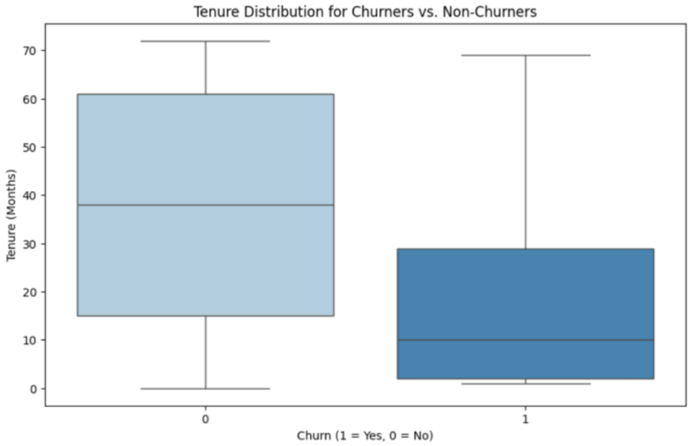
From the ANN model predictions and feature analysis, the following factors were identified:

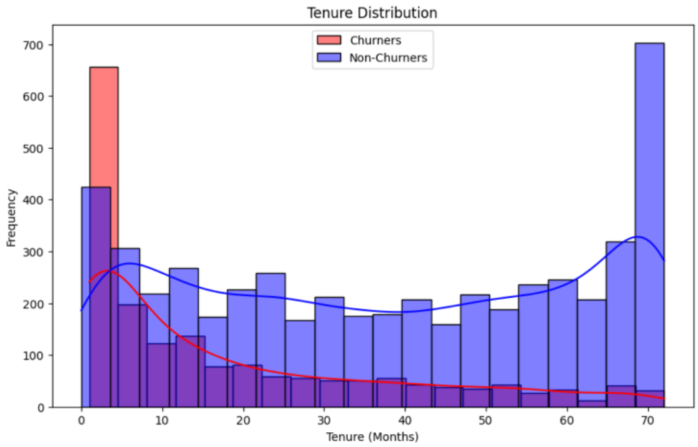
1. **Tenure**

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Compare tenure distribution for churners and non-churners
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x='Churn_Yes', y='tenure', palette='Blues', showfliers=False)
plt.title('Tenure Distribution for Churners vs. Non-Churners')
plt.xlabel('Churn (1 = Yes, 0 = No)')
plt.ylabel('Tenure (Months)')
plt.show()

# Histogram for better distribution insight
plt.figure(figsize=(10, 6))
sns.histplot(data[data['Churn_Yes'] == 1]['tenure'], kde=True, label='Churners', color='red', bins=20)
sns.histplot(data[data['Churn_Yes'] == 0]['tenure'], kde=True, label='Non-Churners', color='blue', bins=20)
plt.title('Tenure Distribution')
plt.xlabel('Tenure (Months)')
plt.ylabel('Frequency')
plt.legend()
plt.show()
```




Churners: Concentrated around shorter tenure values (e.g., <12 months).

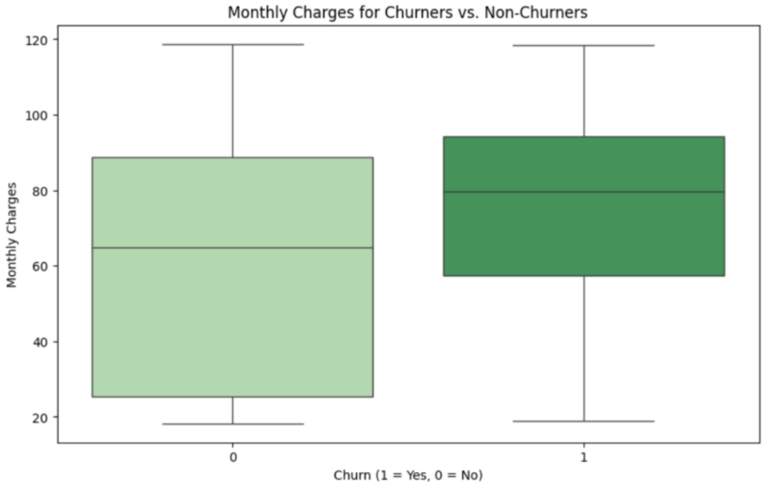
Non-Churners: Show a wider range of tenures, with more customers having longer tenure.

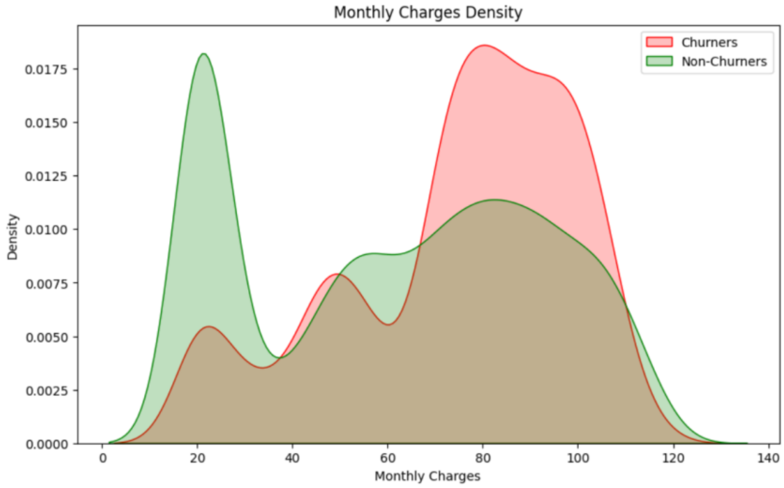
Customers with less than a year of tenure were more likely to churn, indicating dissatisfaction or a failure to engage new customers effectively.

## 2. Monthly charges:

```
# Box plot for monthly charges
plt.figure(figsize=(10, 6))
sns.boxplot(data=data, x='Churn_Yes', y='MonthlyCharges', palette='Greens', showfliers=False)
plt.title('Monthly Charges for Churners vs. Non-Churners')
plt.xlabel('Churn (1 = Yes, 0 = No)')
plt.ylabel('Monthly Charges')
plt.show()

# Kernel density plot
plt.figure(figsize=(10, 6))
sns.kdeplot(data[data['Churn_Yes'] == 1]['MonthlyCharges'], label='Churners', color='red', fill=True)
sns.kdeplot(data[data['Churn_Yes'] == 0]['MonthlyCharges'], label='Non-Churners', color='green', fill=True)
plt.title('Monthly Charges Density')
plt.xlabel('Monthly Charges')
plt.ylabel('Density')
plt.legend()
plt.show()
```




Churners: tend to have higher monthly charges.

Non-Churners: more evenly distributed across lower charge ranges.

Customers with higher monthly charges are more likely to churn

Higher pricing plans correlated with increased churn rates, especially for customers perceiving limited value.

3. **Contract Type:**

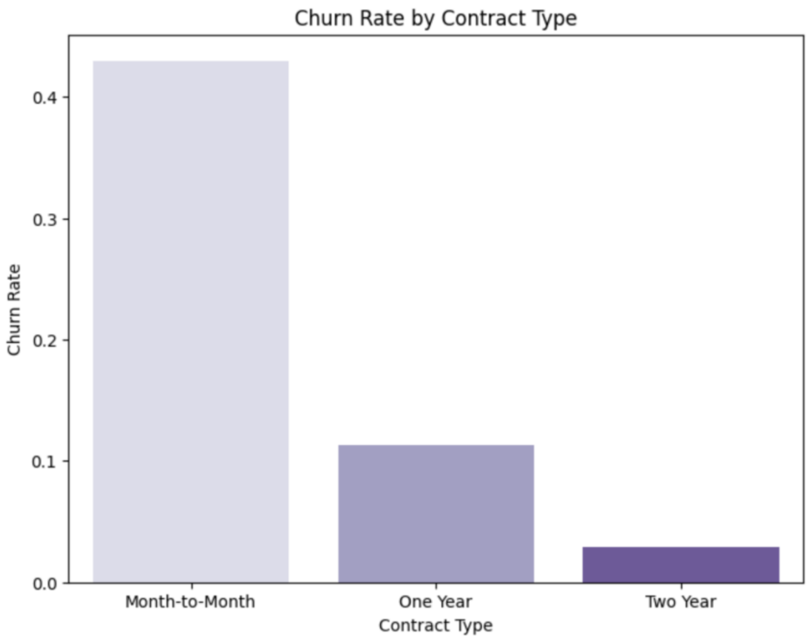
```python
# Create a new 'Contract' column based on one-hot encoded columns
def contract_type(row):
    if row['Contract_One year'] == 1:
        return 'One Year'
    elif row['Contract_Two year'] == 1:
        return 'Two Year'
    else:
        return 'Month-to-Month'

# Apply the function to create the new column
data['Contract'] = data.apply(contract_type, axis=1)

# Calculate churn rates by contract type
contract_churn = data.groupby('Contract')['Churn_Yes'].mean().reset_index()

# Contract type vs churn rate
contract_churn = data.groupby('Contract')['Churn_Yes'].mean().reset_index()

plt.figure(figsize=(8, 6))
sns.barplot(data=contract_churn, x='Contract', y='Churn_Yes', palette='Purples')
plt.title('Churn Rate by Contract Type')
plt.xlabel('Contract Type')
plt.ylabel('Churn Rate')
plt.show()
```




Customers on month-to-month contracts are more likely to churn compared to those on longer-term contracts.

Month-to-Month Contracts: Significantly higher churn rates.

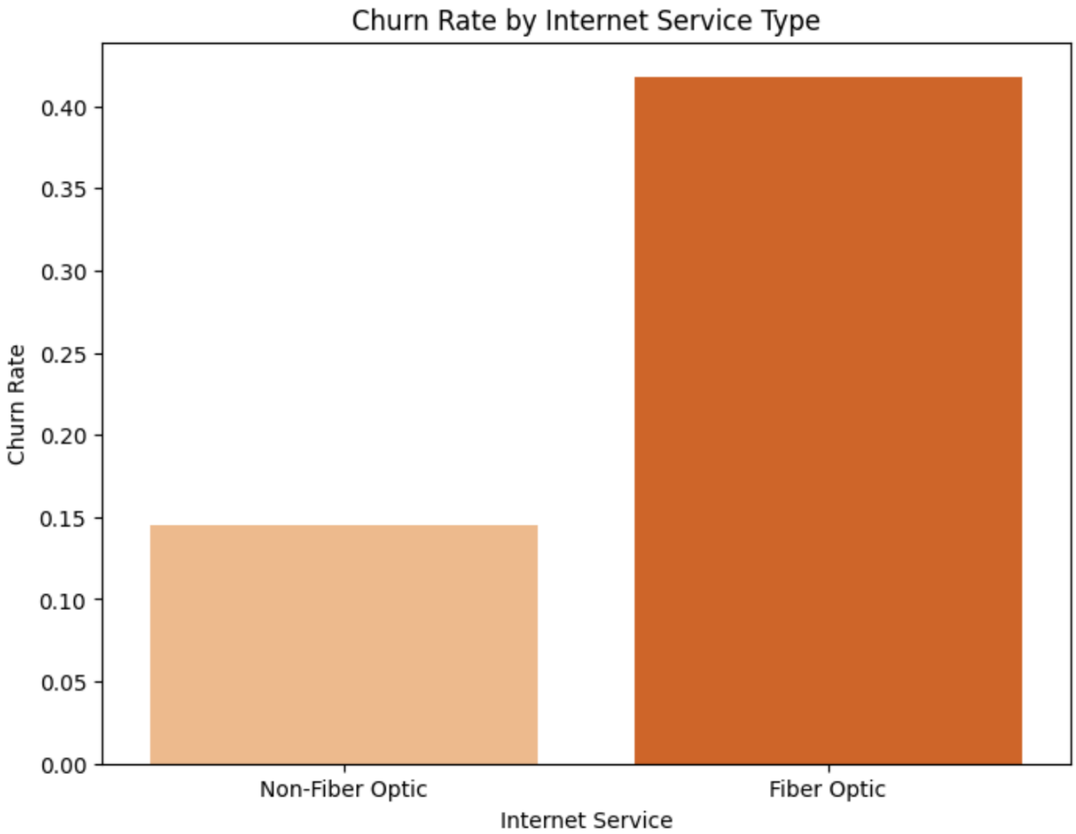
One-Year/Two-Year Contracts: Lower churn rates, indicating greater retention among committed customers.

4. **Service type:**

```python
# Calculate churn rates for customers with and without Fiber Optic service
fiber_churn = data.groupby('InternetService_Fiber optic')['Churn_Yes'].mean().reset_index()

# Rename the column for better readability
fiber_churn['InternetService'] = fiber_churn['InternetService_Fiber optic'].map({0: 'Non-Fiber Optic', 1: 'Fiber Optic'})

# Plot churn rates for Fiber Optic and Non-Fiber Optic services
plt.figure(figsize=(8, 6))
sns.barplot(data=fiber_churn, x='InternetService', y='Churn_Yes', palette='Oranges')
plt.title('Churn Rate by Internet Service Type')
plt.xlabel('Internet Service')
plt.ylabel('Churn Rate')
plt.show()
```




Fiber-optic customers are more likely to churn compared to DSL users

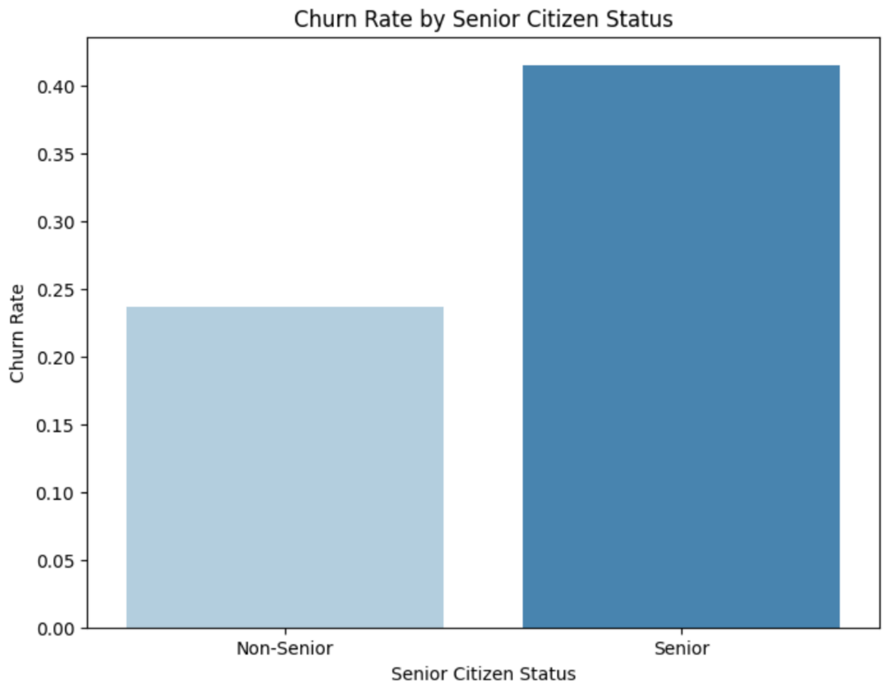
Fiber-Optic Users: Higher churn rates, potentially due to higher costs or unmet expectations.

DSL Users: Lower churn rates, possibly reflecting satisfaction with lower-cost, reliable options.

5. **Senior citizen:**

```
# Churn Rate by Senior Citizen Status
senior_churn = data.groupby('SeniorCitizen')['Churn_Yes'].mean().reset_index()
senior_churn['SeniorCitizen'] = senior_churn['SeniorCitizen'].map({0: 'Non-Senior', 1: 'Senior'})

# Plot the churn rate by senior citizen status
plt.figure(figsize=(8, 6))
sns.barplot(data=senior_churn, x='SeniorCitizen', y='Churn_Yes', palette='Blues')
plt.title('Churn Rate by Senior Citizen Status')
plt.xlabel('Senior Citizen Status')
plt.ylabel('Churn Rate')
plt.show()
```




## Recommendations for Targeted Retention Strategies

| Problems | Possible reasons | Solution |
| --- | --- | --- |
| High churn among customers with short tenure | Due to lack of engagement or dissatisfaction in the early stages of their subscription. | Implement a welcome program to improve the onboarding experience.<br><br>Offer targeted discounts or personalized engagement during the first 6 months.<br><br>Proactively target customers in their first year with offers like free upgrades, bundled services, or reduced rates to extend their tenure.<br><br>Use email or SMS campaigns to engage short-tenure customers with service tips, feedback requests, and problem-solving initiatives.<br><br>Conduct regular surveys in the first six months to identify dissatisfaction early and address it proactively. |

| | | |
|---|---|---|
| High monthly charges increase churn | Customers are cost-sensitive | Introduce tiered pricing plans to cater to cost-sensitive customers.<br><br>Provide transparent billing and emphasize value-added features in communication.<br><br>Provide discounts or flexible payment options for high-paying customers showing signs of dissatisfaction or churn risk.<br><br>Offer periodic discounts or loyalty rewards to customers with high charges who remain subscribed for an extended period. |
| Month-to-month customers are more likely to churn | Due to the flexibility to leave anytime | Offer incentives (e.g., discounts or perks) for customers to switch to longer-term contracts.<br><br>Communicate the benefits of stability and savings with long-term plans<br><br>Address concerns about long-term commitments by including opt-out clauses or partial refunds for one- and two-year contracts.<br><br>For customers nearing the end of their contracts, provide attractive renewal offers to reduce the likelihood of churn. |
| Fiber-optic customers show higher churn rates | Due to higher costs or unmet performance expectations. | Monitor service quality for fiber-optic plans and proactively address complaints.<br><br>Highlight the advantages of fiber-optic services in marketing campaigns to reduce perception gaps.<br><br>Offer exclusive features for Fiber Optic users, such as: streaming service bundles, enhanced |

| | | |
|---|---|---|
| | | customer support, free router or hardware upgrades. |
| Senior citizens' churn rates are higher | Due to cost sensitivity, limited digital adoption, or dissatisfaction with high-cost services like Fiber Optic internet. | Design tailored plans with lower monthly charges or discounted bundles specifically for senior customers.<br><br>Provide senior citizens with personalized support services, such as: dedicated customer service lines for seniors or in-home/remote technical support to resolve issues promptly.<br><br>Simplify billing explanations for senior customers who may be less comfortable with digital payment systems. |

## Documentation of Limitations and Proposed Solutions

**Class imbalance:**

The dataset had significantly more non-churners than churners, creating a class imbalance problem. This skewed the model's performance, favouring the majority class (non-churners) and resulting in lower recall for the minority class (churners).

Impact:

- The model struggled to identify actual churners, as indicated by a lower recall score for the churn class.
- Precision and F1-scores for the churn class were also affected, reducing the model's ability to prioritize at-risk customers.

Solution:

- Generate synthetic examples for the minority class using techniques like SMOTE (Synthetic Minority Oversampling Technique).
- Reduce the number of samples in the majority class to create a balanced dataset.
- Assign higher weights to churners during training to penalize misclassifications of the minority class more heavily.

**Limited feature set:**

The dataset lacked key customer-centric features, such as:

- Satisfaction scores: insights into customer satisfaction levels.
- Service feedback: qualitative data from surveys or support tickets.

- Competitor influence: information about market trends or competitors.

These missing features could provide critical insights into why customers churn or remain loyal.

Impact:

- The model's ability to explain churn behaviour was limited, leading to incomplete insights.

Solution:

- Incorporate data such as:
  - Customer satisfaction: Use surveys or Net Promoter Scores (NPS).
  - Support tickets: Analyse complaints or unresolved issues.
  - Market trends: Add external data on competitor pricing and promotions.
- Create new features from existing data, such as:
  - Tenure in specific plans.
  - Average monthly charges over time.

**Static predictions:**

The model used historical data to predict churn, but it didn't account for real-time changes in customer behaviour, such as:

- Recent complaints or support tickets.
- Sudden changes in usage patterns or billing activity.

Impact:

- The predictions might not reflect the current churn risk, reducing the relevance of retention strategies.

Solution:

- Develop a **real-time churn prediction system** by integrating streaming data into the model pipeline.
- Include triggers for events like:
  - Sudden billing changes.
  - Negative sentiment in support tickets.
  - Usage drops or complaints.

**Suboptimal threshold:**

The default threshold of 0.5 for classification might not align with business goals, such as prioritizing recall (identifying more churners) over precision.

Impact:

- The model's predictions didn't fully meet the business requirement of reducing missed churners (false negatives).

Solution:

- Conduct threshold optimization to find the value that aligns with business goals:
  - Use precision-recall curves to determine the threshold that balances recall and precision.